The Digital Library Landscape

Looking for Trends

William Y. Arms
Department of Computer Science
Cornell University

CORNELL

Primary Information

CORNELL

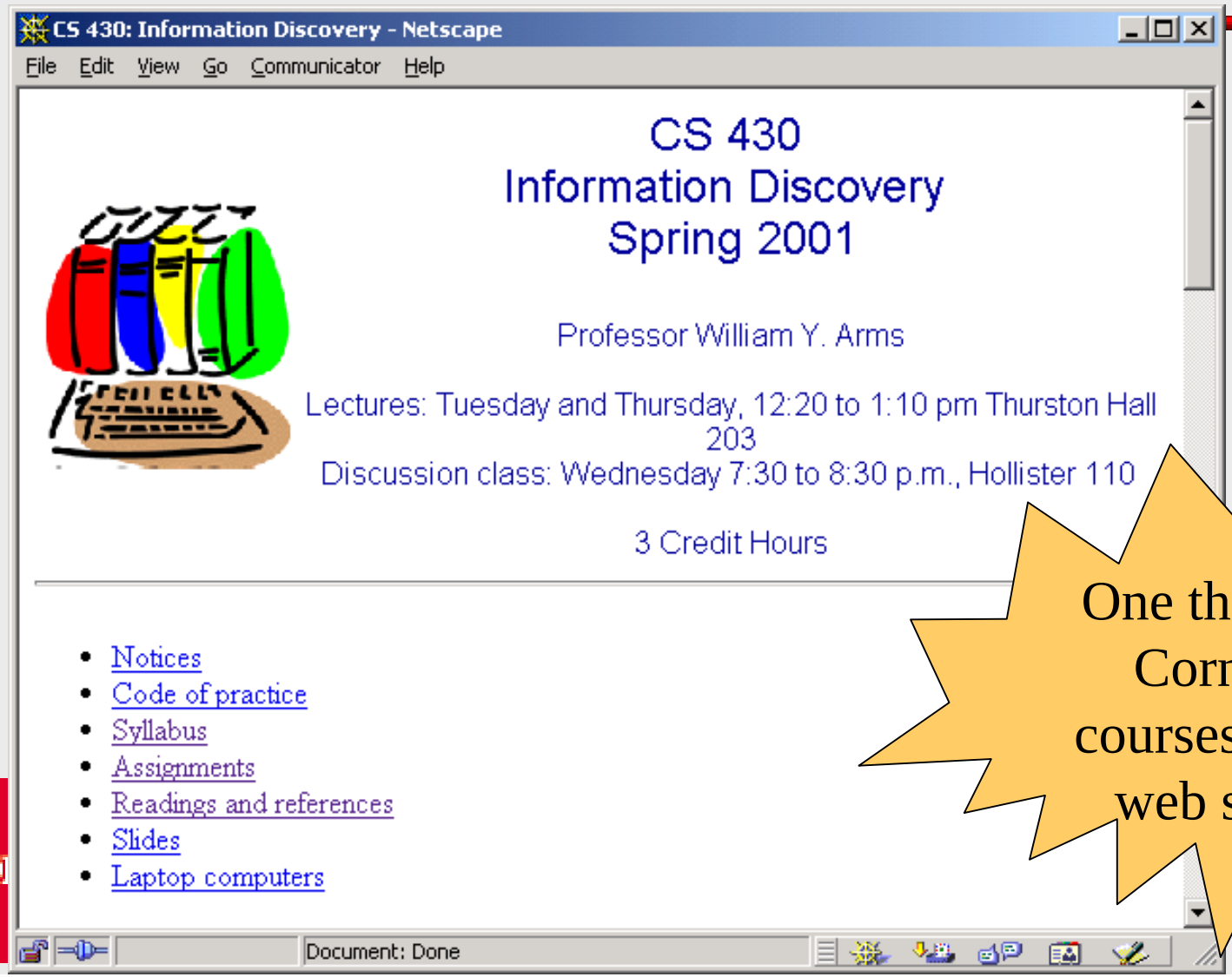
Underlying Trends

Every year sees an increase in the proportion of important information that is available with open access.

Every year sees an increase in the proportion of important information that is available online.

CORNELL

Course Web Sites



One third of Cornell courses have web sites

MIT News

MIT to make nearly all course materials available free on the World Wide Web

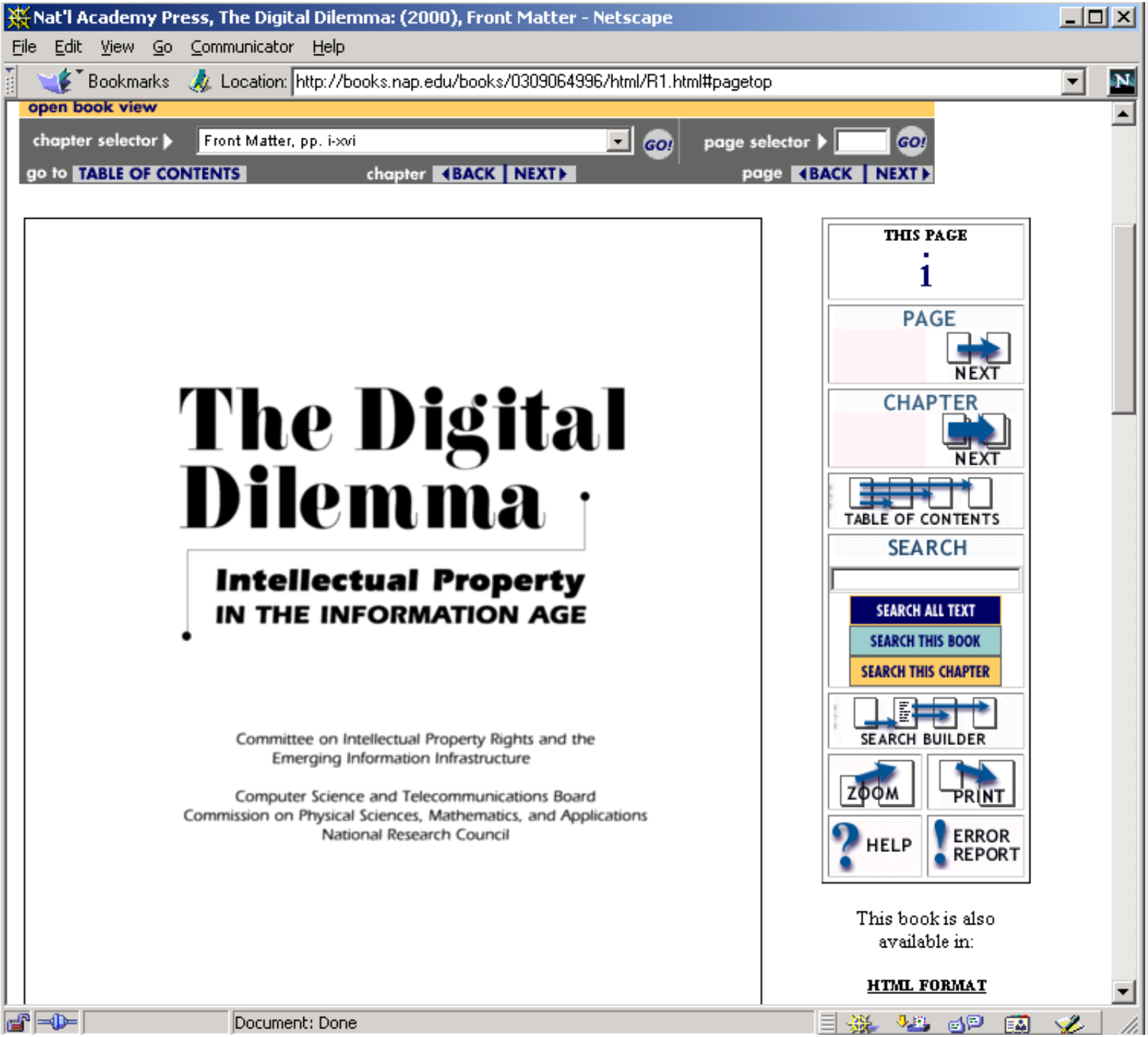
Unprecedented step challenges 'privatization of knowledge'

CAMBRIDGE, Mass. -- MIT President Charles M. Vest has announced that the Massachusetts Institute of Technology will make the materials for nearly all its courses freely available on the Internet over the next ten years. He made the announcement about the new program, known as MIT OpenCourseWare (MITOCW), at a press conference at MIT on Wednesday, April 4th.

MIT Press Release, April 4, 2001

CORNELL


arXiv.org e-Print archive - Netscape
File Edit View Go Communicator Help
Bookmarks Location: http://xxx.lanl.gov/
arXiv.org e-Print archive
Automated e-print archives at arXiv.org: physics Form Interface .
6 Jan '00: Cumulative "What's New" pages
Guide to the perplexed: Read these PROFESSIONAL HELP pages before sending fan mail.
ROBOTS BEWARE: indiscriminate automated downloads from this site are not permitted.
Physics
Astrophysics (astro-ph new, recent, abs, find)
Condensed Matter (cond-mat new, recent, abs, find)
includes: Disordered Systems and Neural Networks; Materials Science; Mesoscopic Systems and Quantum Hall Effect; Soft Condensed Matter; Statistical Mechanics; Strongly Correlated Electrons; Superconductivity
General Relativity and Quantum Cosmology (gr-qc new, recent, abs, find)
High Energy Physics - Experiment (hep-ex new, recent, abs, find)
High Energy Physics - Lattice (hep-lat new, recent, abs, find)
High Energy Physics - Phenomenology (hep-ph new, recent, abs, find)
High Energy Physics - Theory (hep-th new, recent, abs, find)
Mathematical Physics (math-ph new, recent, abs, find)
Nuclear Experiment (nucl-ex new, recent, abs, find)
Document: Done


Nat'l Academy Press, The Digital Dilemma: (2000), Front Matter - Netscape
File Edit View Go Communicator Help
Bookmarks Location: http://books.nap.edu/books/0309064996/html/R1.html#pagetop
open book view
chapter selector Front Matter, pp. i-xvi GO!
page selector GO!
go to TABLE OF CONTENTS
chapter BACK NEXT
page BACK NEXT
The Digital Dilemma
Intellectual Property
IN THE INFORMATION AGE
Committee on Intellectual Property Rights and the
Emerging Information Infrastructure
Computer Science and Telecommunications Board
Commission on Physical Sciences, Mathematics, and Applications
National Research Council
THIS PAGE
i
PAGE
NEXT
CHAPTER
NEXT
TABLE OF CONTENTS
SEARCH
SEARCH ALL TEXT
SEARCH THIS BOOK
SEARCH THIS CHAPTER
SEARCH BUILDER
ZOOM
PRINT
HELP
ERROR REPORT
This book is also available in:
HTML FORMAT
Document: Done

The New York Times on the Web - Netscape

File Edit View Go Communicator Help

Bookmarks Netsite: http://www.nytimes.com/

GET HOME DELIVERY

The New York Times ON THE WEB

CLICK HERE education A SPECIAL REPORT life

Login/Register | SATURDAY, FEBRUARY 3, 2001 5:44 PM ET | Personalize Your Weather

SEARCH

Latest News
Archives
Search Tips

Jobs
Real Estate
Shopping

NEWS

Business
Editorial/Op-Ed
International
National
New York Region
NYT Front Page
Obituaries
Politics
Quick News
Health
Science
Sports
Tech/Internet
Weather

Corrections

FEATURES

Arts
Automobiles
Books
Cartoons
Crossword

The Making of a Suspect: The Case of Wen Ho Lee

By MATTHEW PURDY
FROM SUNDAY'S TIMES
A re-examination of the case against Los Alamos scientist Wen Ho Lee indicates that investigators took ambiguous evidence about Dr. Lee's behavior and Chinese atomic espionage and wove it into a case that eventually collapsed. The first of a two-part series. Go to Article

▪ DIAGRAM: Building a Smaller H-Bomb

U.S. Tries Defusing Allies' Opposition to Missile Defense

By MICHAEL R. GORDON
FROM SUNDAY'S TIMES
Defense Secretary Donald H. Rumsfeld, the first senior Bush administration official to visit Europe, tried today to defuse opposition to the administration's antimissile plans by offering to help European nations and other allies to deploy missile defenses.

(NYT)
After a five-year investigation, Dr. Wen Ho Lee, pictured here in December, was all but threatened with execution for not admitting spying. But prosecutors were never able to connect him to espionage. Go to Article

NATIONAL

Field Guide? Check. Binoculars? Check. Lobbyists? Soon. (12:00 a.m.)

INTERNATIONAL

Flames Consume the Dead, but Not the Anguish of India (12:00 a.m.)

INSIDE More

TRAVEL

Spring-Summe Cruise Special

Tahiti, Antigua, Murmansk.
Cruise ships set sail for distant ports. Also in Travel: Deal of the Day.

ABUZZ

Stocking the Shelves?

CLG815 asks: What are the twenty most essential items to have in the pantry? Add your suggestions.

discuss
ask
answer
review

MARKETS At Clos

Dow	10864.10	-119.53
Nasdaq	2660.50	-122.29
S&P 500	1349.47	-24.00
Russell 2000	501.50	-7.44

Document: Done

D-Lib Magazine

ISSN: 1082-9873

Vol. 7 No. 1

January 2001

EDITORIAL

BOOK REVIEW

"...This book about cataloging certainly deserves to be read outside the community whose principles and traditions it describes and illuminates."

Implements Used on the Farm: Threshing Machine (Northern Great Plains) - Netscape

File Edit View Go Communicator Help

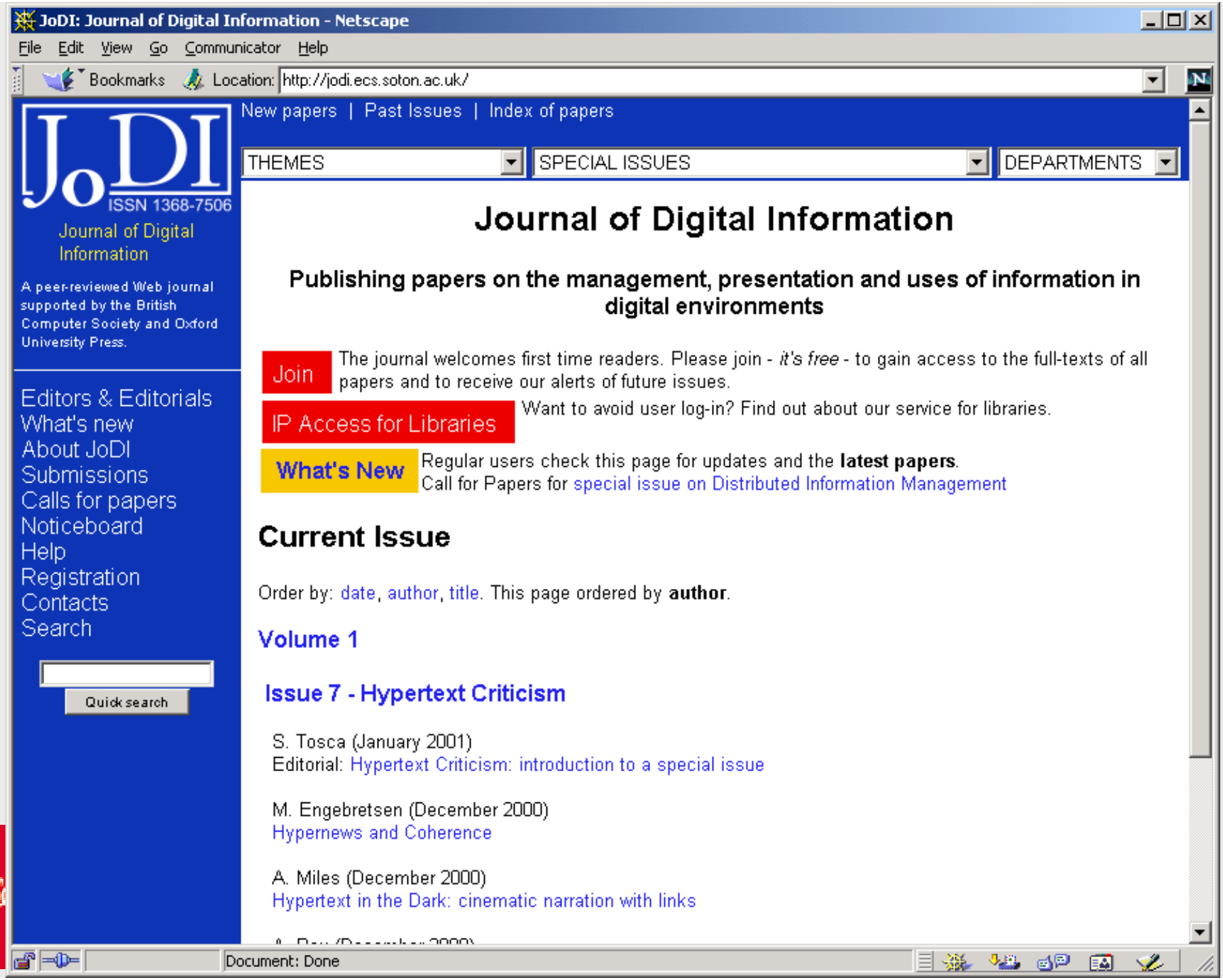
Bookmarks Location: http://memory.loc.gov/ammem/award97/ndfahtml/ngp_farm_threshing.html

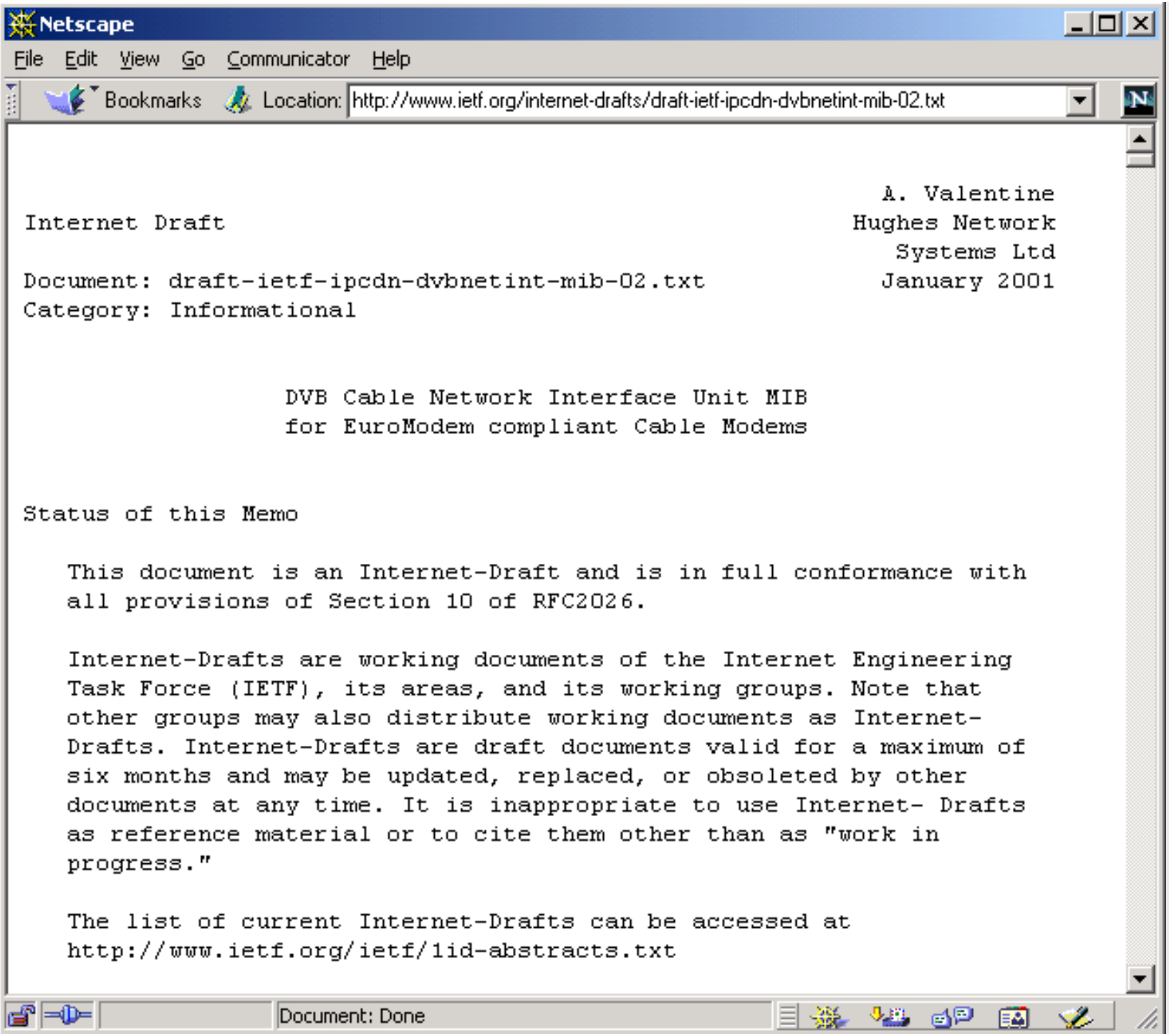
Northern Great Plains, 1880-1920

Search on: Threshing | Threshing Machines | Threshing Crews

Document: Done


National Weather Service Home Page -U.S. Government warnings & forecasts, Organization, Online Docum - Nets...
File Edit View Go Communicator Help
Bookmarks Location: http://www.nws.noaa.gov/
National Oceanic and Atmospheric Administration
National Weather Service Homepage
[Organization & Services | Mission Statement | Strategic Plan]
Espanol
online
Direct Access to U.S.
OFFICIAL WEATHER FORECAST PRODUCTS & OBSERVATIONS
Search
this site
NATIONAL
Weather Data
Warnings
Forecast Products
Current Conditions
Weather Maps
Storm Prediction
Marine Weather
Extended Forecasts
INTERNATIONAL
LATEST RADAR IMAGE
has Site Radar Links
Internet Weather Source
SELECT A STATE GRAPHICAL
LATEST RADAR LOOP
10 frame animation
Interactive Weather Information Network
IWIN
HIGHLIGHTED PRODUCTS
Storm Prediction Center's
Watch/Warning Display
(20 minute update)
Hydrometeorological Prediction Center's
24 Hour Surface Forecast Chart
(twice daily)
National Weather Service REGIONS
Web Sites (map)
Alaska
Central
Eastern
Pacific
Southern
Western
SPECIALIZED CENTERS
Document: Done


JoDI: Journal of Digital Information - Netscape
File Edit View Go Communicator Help
Bookmarks Location: http://jodi.ecs.soton.ac.uk/
JoDI
ISSN 1368-7506
Journal of Digital Information
A peer-reviewed Web journal supported by the British Computer Society and Oxford University Press.
New papers | Past Issues | Index of papers
THEMES
SPECIAL ISSUES
DEPARTMENTS
Editors & Editorials
What's new
About JoDI
Submissions
Calls for papers
Noticeboard
Help
Registration
Contacts
Search
Quick search
Journal of Digital Information
Publishing papers on the management, presentation and uses of information in digital environments
Join
The journal welcomes first time readers. Please join - it's free - to gain access to the full-texts of all papers and to receive our alerts of future issues.
IP Access for Libraries
Want to avoid user log-in? Find out about our service for libraries.
What's New
Regular users check this page for updates and the latest papers.
Call for Papers for special issue on Distributed Information Management
Current Issue
Order by: date, author, title. This page ordered by author.
Volume 1
Issue 7 - Hypertext Criticism
S. Tosca (January 2001)
Editorial: Hypertext Criticism: introduction to a special issue
M. Engebretsen (December 2000)
Hypernews and Coherence
A. Miles (December 2000)
Hypertext in the Dark: cinematic narration with links
Document: Done


Netscape
File Edit View Go Communicator Help
Bookmarks Location: http://www.ietf.org/internet-drafts/draft-ietf-ipcdn-dvbnetint-mib-02.txt
Internet Draft
A. Valentine
Hughes Network
Systems Ltd
January 2001
Document: draft-ietf-ipcdn-dvbnetint-mib-02.txt
Category: Informational
DVB Cable Network Interface Unit MIB
for EuroModem compliant Cable Modems
Status of this Memo
This document is an Internet-Draft and is in full conformance with all provisions of Section 10 of RFC2026.
Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts. Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet- Drafts as reference material or to cite them other than as "work in progress."
The list of current Internet-Drafts can be accessed at
http://www.ietf.org/ietf/1id-abstracts.txt
Document: Done

Public Library of Science

Open Letter

We support the establishment of an online public library that would provide the full contents of the published record of research and scholarly discourse in medicine and the life sciences in a freely accessible, fully searchable, interlinked form. Establishment of this public library would vastly increase the accessibility and utility of the scientific literature, enhance scientific productivity, and catalyze integration of the disparate communities of knowledge and ideas in biomedical sciences.

CORNELL

Secondary Information

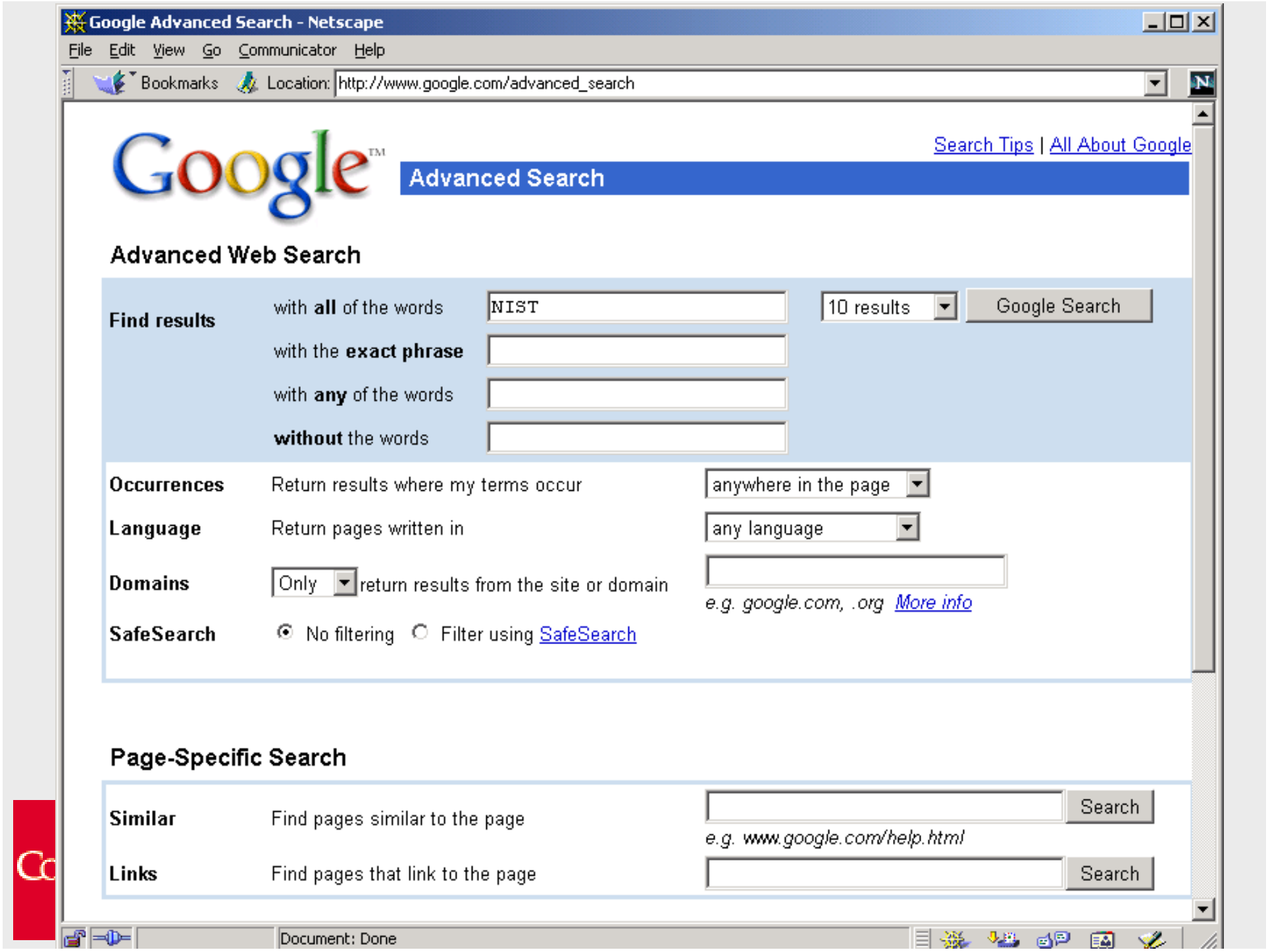
CORNELL

Information Discovery

"I used to be a heavy user of Inspec. Now I use Google instead."

Why are web search services the most widely used information discovery tools in universities today?

CORNELL


Google Advanced Search - Netscape
File Edit View Go Communicator Help
Bookmarks Location: http://www.google.com/advanced_search
Search Tips | All About Google
Google
Advanced Search
Advanced Web Search
Find results
with all of the words
NIST
10 results
Google Search
with the exact phrase
with any of the words
without the words
Occurrences
Return results where my terms occur
anywhere in the page
Language
Return pages written in
any language
Domains
Only
return results from the site or domain
e.g. google.com, .org More info
SafeSearch
No filtering
Filter using SafeSearch
Page-Specific Search
Similar
Find pages similar to the page
Search
e.g. www.google.com/help.html
Links
Find pages that link to the page
Search
Document: Done

Amazon.com: buying info: Digital Libraries (Digital Libraries and Electronic Publishing) - Netscape

File Edit View Go Communicator Help

Bookmarks Location: azon.com/exec/obidos/ASIN/0262011808/qid=981304408/sr=2-1/ref=sc_b_1/107-8463609-4289304

Rate it

Visit the New Cars Store

2001 Jetta

2001 Jetta by Volkswagen

Visit the Music Store

All That You Can't Leave Behind

~ U2

Editorial Reviews

Amazon.com

These days we seem to be creating information faster than we can store it, but the near future is looking bright. Cornell professor William Y. Arms offers a program for that future in *Digital Libraries*, a synthesis of library and computer sciences that presents the history and current developments in each field with special emphasis on their interactions.

Since the book necessarily must appeal to a broad spectrum of professionals, any given reader will find some parts elementary, but Arms clearly maps the common ground and much of the text will appeal to all. Chapters covering the basics of information management, the Internet, security, archives, and retrieval bridge the traditional books-and-shelves library systems and the often jury-rigged information architecture developed over 40 years of computer use.

Digital Libraries contains plenty of sidebars detailing historical information as well as definitions primarily suited to professionals entering the interdisciplinary zone (but which would unacceptably break up the text flow; while it's important to understand both MARC codes and TCP/IP protocols, it's best for each reader to decide what supplementary information is needed). *Digital Libraries* is an ambitious and important book--if we are to develop truly efficient and accessible information management systems, everyone concerned must understand their shared technical history and move forward as one. --*Rob Lightner*

Was this content helpful to you? YES NO

Document: Done

Epinions.com - rbinck's profile - Netscape

File Edit View Go Communicator Help

Bookmarks Location: http://www.epinions.com/user-rbinck

Epinions.com℠

Before you buy anything
make an informed decision.

Search

Web of Trust

rbinck trusts:

1. hikini
2. tekki
3. alanrus
4. bmcnichol
5. ptiemann

» View all 62 members whom **rbinck trusts**

rbinck is trusted by:

1. Daxman
2. LordBalfor
3. j_hipps
4. splace
5. gd134

» View all 25 members who **trust rbinck**

rbinck's Profile

Member: **Richard Binckley**
Epinions.com ID: **rbinck**
Location: **Houston, Texas - Kailua-Kona, Hawaii**
Member since: **Dec 27 '00**

Opinions written: **58**
Products rated: **63**
Member visits: **-**
Total visits: **-**

Part time Hawaii resident that likes Hawaii and Classic Cars.

rbinck's Recent Opinions

Date Written	Opinion Title	Product / Topic	Product Rating	Opinion Rating
Mar 15 '01	**Kealakekua Bay - Snorkeling at it's Finest**	Kealakekua Bay State Historical Park in Parks	★★★★★	Very Helpful

Document: Done

Before You Ask ...

- The open access information is sometimes a poor substitute
- Much good information is not available with open access

CORNELL

Economics

CORNELL

The Dilemma

It is hard to compete with a free good.

Library budgets and publishers' revenues are vulnerable.

Yet money is needed to pay for professional staff.

CORNELL

Four Economic Models

Example: Broadcast Television

Open Access

Advertising	network television
External funding	public broadcasting

Restricted Access

Subscription	cable
Pay-by-use	pay-per-view

CORNELL

Examples

Old	New
Books in Print (*subscription*)	Amazon.com (*advertising*)
Medline (*pay-by-use*)	Grateful Med (*external*)
Journal (*subscription*)	ePrint archives (*external*)
Westlaw (*pay-by-use*)	Legal Information Institute (*external*)
Inspec (*subscription*)	Google (*advertising*)

CORNELL

A False Assumption

Incorrect thinking

The only incentive for creating information is to make money -- royalties to authors and profits for publishers

Correct thinking

Many creators do not require revenue

- Marketing and promotion
- Government information
- Academic research

They want their materials to be used

CORNELL

Scholarly Information

The dominant force is **author pressure**, which emphasizes **open access** rather than closed access.

1. A mixture of economic models will coexist.

2. Eventually, we will have open access to most scientific, government and professional information.

3. The most common economic model will be that information is published by the producing organization.

CORNELL

The Cost of Libraries and Publishing

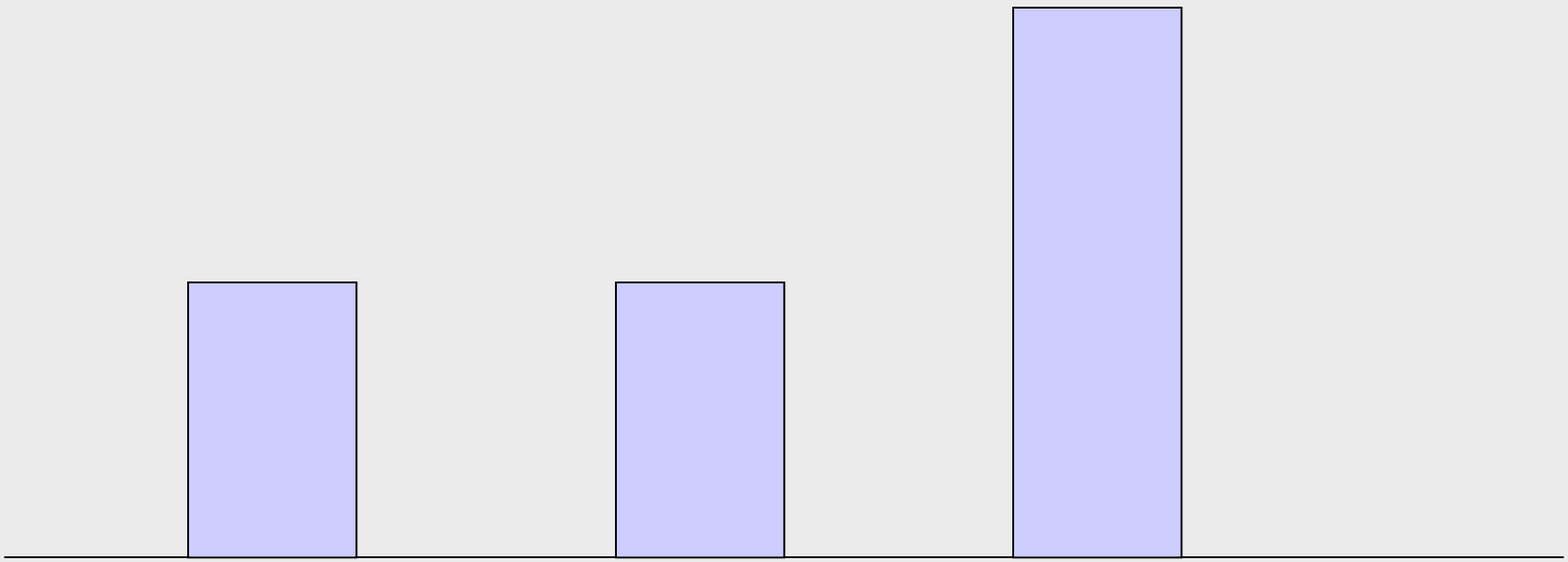
The costs of libraries and publishing are dominated by personnel.

Major reductions in unit costs require different use of personnel.

By creative use of technology, can we build libraries that are of high quality at much lower costs?

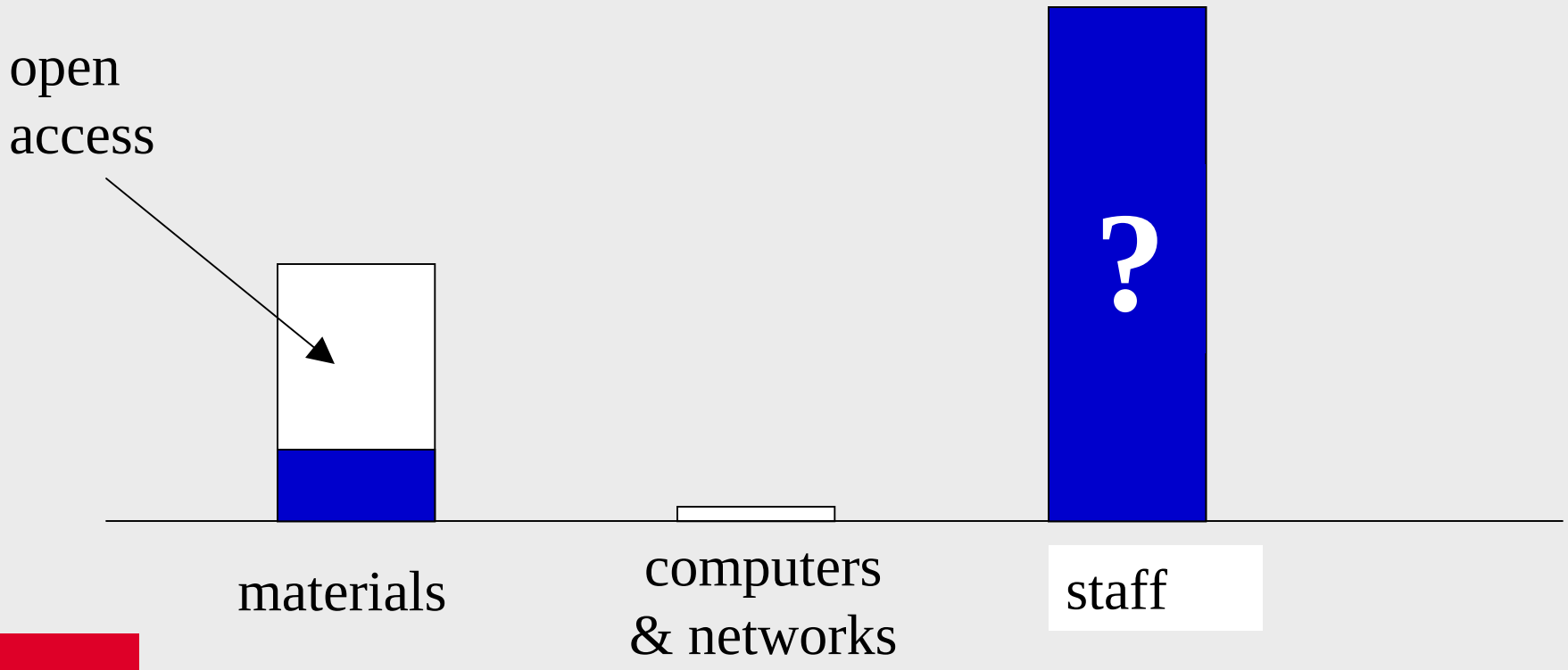
CORNELL

Research Libraries are Expensive



CORNELL

The Potential of Digital Libraries



CORNELL

Dramatic Reductions in Cost

Thought experiment: *How would you reduce the cost of scientific, legal, medical and government information to one fifth?*

The only possible answer: *Automate labor intensive tasks.*

Moore's Law is the only hope.

CORNELL

Brute Force Computing

Few people really understand Moore's Law

-- Computing power doubles every 18 months

-- Increases 100 times in 10 years

-- Increases 10,000 times in 20 years

Simple algorithms + immense computing power may outperform human intelligence.

CORNELL

Automated Digital Libraries: Examples

Automatic indexing	Lycos, Infoseek, Altavista, Google, ...
Query matching	Vector methods (Salton)
Ranking importance	Google (Page and Brin)
Archiving	Internet Archive (Kahle)
Collection development	ResearchIndex (Lawrence)
Metadata extraction	Informedia (Wactlar)

Example: Catalogs and Indexes

Catalog, index and abstracting records are very expensive when created by skilled professionals, *but* ...

For information discovery, particularly with untrained users:

automated indexing of full text

is at least as effective as

manually produced indexes and catalogs

[Demonstrated repeatedly in experiments going back to the original Cranfield experiments.]

CORNELL

The National Science Library (NSDL)

Can we build a very low cost national science library -- initially for education -- using the methods of automated digital libraries?

CORNELL

Site for science

One of Six Core Integration Demonstration Projects for the NSDL

CORNELL

How Big might the NSDL be?

The NSDL aims to be comprehensive -- all branches of science, all levels of education, very broadly defined.

Five year targets:

1,000,000	different users
10,000,000	digital objects
100,000	independent sites

Requires: **low-cost, scalable, technology**
automated collection building and maintenance

Site for science

The Spectrum of Interoperability: Federation

Standardization on sophisticated protocols, formats, metadata, authentication, etc.

Examples:

Library catalogs with MARC and Z 39.50

DLESE (NSDL)

smete.org (NSDL)

- High-quality interoperability of services
- High cost of entry to participating sites

<u>Smallish numbers of tightly integrated partners</u>

Has difficulty scaling

Site for science

The Spectrum of Interoperability: Metadata Harvesting

Agreements on simple protocol and metadata standard(s)

Example:

Metadata harvesting protocol of
the Open Archives Initiative (MHP)

- Moderate-quality services
- Low cost of entry to participating sites

Moderately large numbers of loosely collaborating sites

Promising but still an emerging approach

The Spectrum of Interoperability: Gathering

Robots gather collections automatically with no participation from individual sites

Examples:

Web search services (e.g., Google)

CiteSeer (a.k.a. ResearchIndex)

- Restricted but useful services
- Zero cost of entry to gathered sites

Very large numbers of independent sites

Only suitable for open access collections

Federal Agencies

How can the federal agencies help?

CORNELL

As a Supplier of Information

Primary information

- Online, preferably with open access
- Support the interoperability spectrum, (e.g., the Metadata Harvesting Protocol of the Open Archives Initiative)

Secondary information

- Online, preferably with open access

CORNELL

The Open Access Web

Before the web

- Few people had access to scientific, medical, government and legal information

With the web

- Much high quality information is available with open access
- Low cost services can organize this information and provide open access to it

"Please can I use the web? I don't do libraries."
Anonymous Cornell student, circa 1996.

CORNELL

Some Light Reading

William Y. Arms, "Automated digital libraries." *D-Lib Magazine*, July/August 2000.
http://www.dlib.org/dlib/july20/07contents.html

William Y. Arms, "Economic models for open-access publishing." *iMP*, March 2000.
http://www.cisp.org/imp/march_2000/03_00arms.htm

CORNELL